



Teoría de Información:

Información Mutua.

- Cuando los mensajes que se envían son completamente independientes de los mensajes que se reciben ($p(x,y)=p(x)p(y)$), entonces la $MI(x,y)=0$. Por tanto, si MI es distinta de cero refleja que hay cierta correlación entre los eventos de x e y .
- Para señales continuas se cambian los sumatorios por integrales.
- Otra forma útil de definir la información mutua es a través de entropías, $MI(x,y)=S(x)+S(y)-S(x,y)$. (demostrar para casa (3))
- En el caso de sucesos independientes $S(x,y)=S(x)+S(y)$, y por tanto $MI(x,y)=0$.

Teoría de Información:

Información Mutua: algunas propiedades

- Suponiendo:

$$S(X) = -\sum_i p(x_i) \log_2 (p(x_i)),$$

$$S(X, Y) = -\sum_{xy} p(x, y) \log_2 (p(x, y)),$$

$$S(X|Y) = -\sum_y p(y) \sum_x p(x|y) \log_2 (p(x|y)).$$

- Se puede demostrar:

$$S(X, Y) = S(X) + S(Y|X) = S(Y) + S(X|Y)$$

$$MI(X, Y) = S(X) - S(X|Y) = S(Y) - S(Y|X)$$

$$MI(X, X) = S(X)$$

(demostrar para casa (4))



Teoría de Información:

Información Mutua.

- Supongamos x como estímulo (S) y la variable y como respuesta (R).
- ¿Qué significa que MI_{RS} sea grande o pequeña?
- La entropía es básicamente una medida del desorden, y nos da una idea de la incertidumbre de un conjunto de posibles medidas.
- Supongamos que $MI_{RS}=a$ y $MI_{R\bar{S}}=\underline{a}$, habiendo obtenido como resultado que $a > \underline{a}$.
- Esto significa que la incertidumbre del conjunto R dado el conocimiento de \bar{S} es mayor que la incertidumbre de R dado que conozco S .
- La variabilidad de todas las posibles medidas se reduce a través de la interacción con S , y se reduce menos con el conocimiento de \bar{S} .



Teoría de Información:

Información Mutua.

- Asumimos que la información de la red está completamente contenida en la evolución temporal de la membrana de potencial de las neuronas.
- Vamos a hacer un tratamiento discreto de la señal.
- Hay muchas evidencias experimentales que el procesamiento de información que realizan estos sistemas está gobernado por los eventos discretos que son los spikes y los bursts.
- Muchos autores ya han usado este tratamiento discreto de la señal que forma la membrana de potencial. Para discretizar la evolución temporal de la membrana de potencial dividimos nuestro espacio de tiempo en N ventanas de tiempo de tamaño Δt .



Teoría de Información:

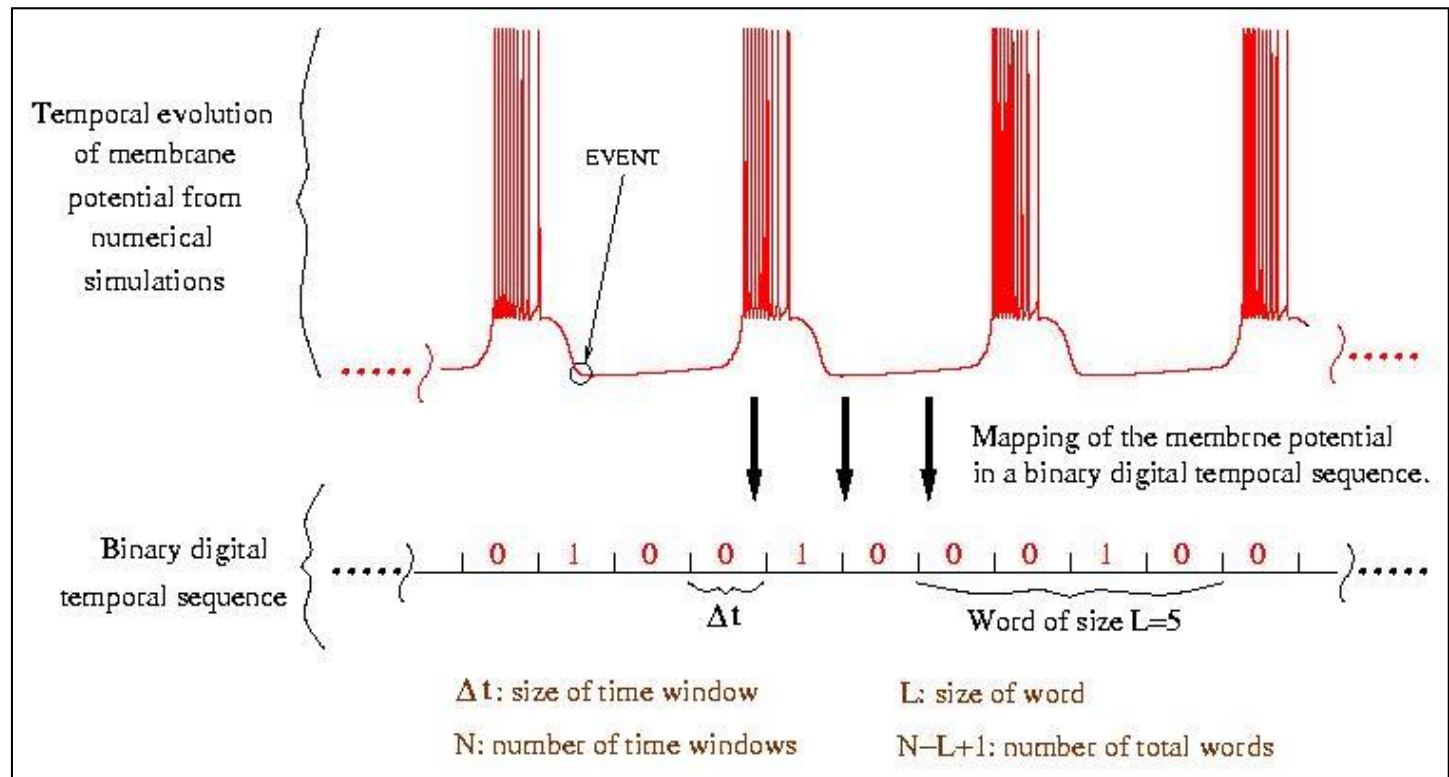
Información Mutua.

- Mapeamos las N ventanas con series temporales de eventos.
- Asignamos el valor de 1 a cada caja de tamaño Δt si tenemos un evento y 0 en cualquier otro caso.
- Suponemos que la información principal de las series esta contenida en la ocurrencia de bursts.
- Hay varios experimentos que corroboran este hecho, por ejemplo en el CPG del ganglio estomatogástrico de la langosta, la frecuencia de la la onda lenta esta íntimamente relacionada con la frecuencia de los músculos del píloro o del estómago.
- Para especificar la ocurrencia de burst buscamos en las regiones de hiperpolarización en las series temporales del potencial de membrana.

Teoría de Información:

Información Mutua.

- Así para codificación de bursts cuando encontramos máxima hiperpolarización asignamos el valor del evento de 1 para la caja donde lo encontremos.





Teoría de Información:

Información Mutua.

- Por lo tanto tenemos una secuencia discreta de eventos temporales $\{e_t, t=1 \dots N\}$ donde e_t puede ser 0 o 1.
- Definimos una palabra W_t^L de tamaño L en el tiempo t , como la secuencia de símbolos $W_t^L = \{e_t, e_{t+1}, \dots, e_{t+L}\}$.
- Por lo tanto tenemos $N-L+1$ palabras de longitud L bits, en cada instante de tiempo.
- Ya tenemos definido nuestro espacio de códigos para medir el intercambio de información entre neuronas.



Teoría de Información:

Información Mutua.

- Para medir la MI consideramos en la red una neurona como salida (R , neurona respuesta), y otra como neurona de entrada.
- Entonces calculamos $P_{RS}(r_i, s_i)$ para estimar el promedio de la información para una secuencia dada de palabras formadas por la respuesta a un estímulo dado.
- Para estimar esta distribución de probabilidad hacemos uso de los datos experimentales o de simulaciones que tengamos.



Teoría de Información:

Información Mutua.

- La medida de información se puede normalizar utilizando la información mutua normalizada o entropía normalizada:

$$E_{RS} = \frac{I_{RS}}{H(S)}$$

- Esta medida se usa para medir la transferencia de información del estímulo a la neurona respuesta.
- Esta medida se encuentra $0 \leq E_{RS} \leq 1$, debido $H(S)$ es la máxima cantidad de información que se puede transmitir del estímulo a la neurona respuesta.

Teoría de Información:

Información Mutua.

- $E_{RS}=0$ significa que toda la información es perdida, es decir respuesta y estímulo son completamente independientes.
- En el otro extremo, cuando $E_{RS}=1$ significa que hay un perfecto seguimiento del estímulo a la respuesta de la neurona (completa sincronización).
- Este análisis es similar a:
 - F.B. Rodriguez, P. Varona, R. Huerta, M.I. Rabinovich, H.D.I. Abarbanel. 2001. Richer network dynamics of intrinsically non-regular neurons measured through mutual information. [*Lect. Notes Comput. Sc.*, 2084: 490-497.](#)
 - F.B. Rodríguez, Roberto Latorre, Pablo Varona. Characterization of Triphasic Rhythms in Central Pattern Generators (II): Burst Information Analysis. Artificial Neural Networks — ICANN 2002 Volume 2415 of the series [*Lecture Notes in Computer Science* pp 167-173.](#)
 - A. Lareo, C.G. Forlim, R.D. Pinto, P. Varona, F.B. Rodriguez. 2016. Temporal Code-Driven Stimulation: Definition and Application to Electric Fish Signaling. [*Frontiers in Neuroinformatics* 10: 41.](#)

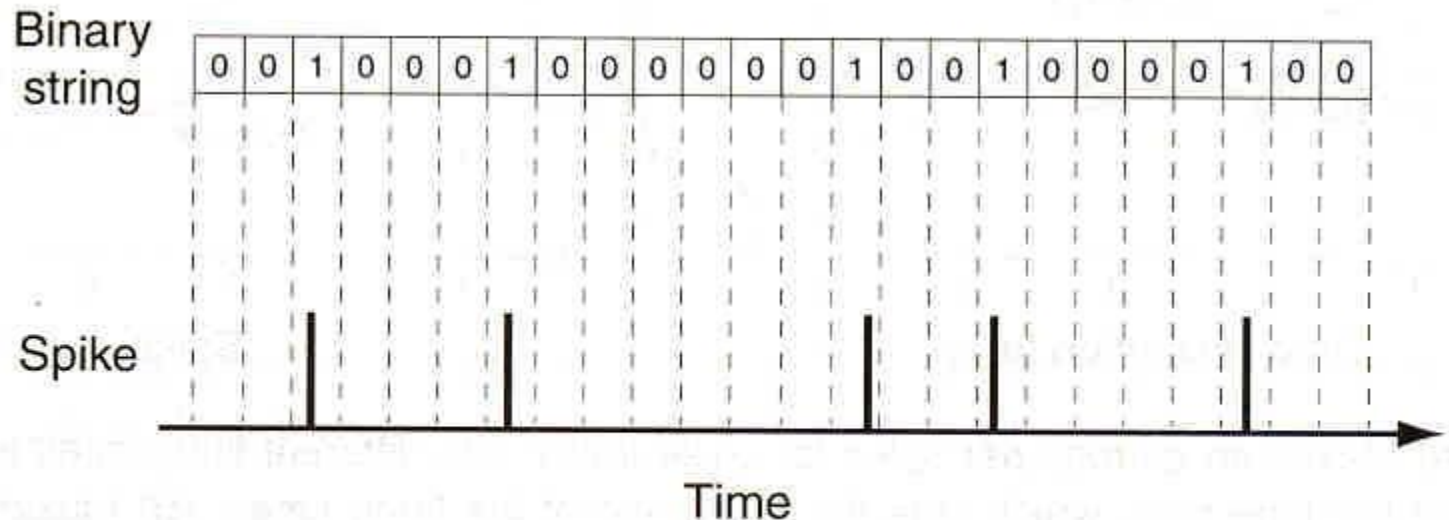
Teoría de Información:

Información Mutua.

- Cualquier método de simbolización de la serie temporal generada por el voltaje de las neuronas es válido.
- Por ejemplo se puede utilizar:
 - Patrones de orden:
 - [C. Bandt, B. Pompe. 2002. Permutation Entropy: A Natural Complexity Measure for Time Series, Phys. Rev. Lett. 88, 174102.](#)
 - [K. Keller, M. Sinn.2005. Ordinal analysis of time series. Physica A: Stat. Mech. Appl. 356, 114.](#)
 - [Bilal Fadlallah, Badong Chen, Andreas Keil, and José Príncipe. Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information. Phys. Rev. E 87, 022911. 2013.](#)
 - [Keller Karsten · Unakafov Anton M. · Unakafova Valentina A. **Ordinal Patterns, Entropy, and EEG.** *Entropy* **2014**, 16\(12\), 6212-6239; doi:10.3390/e16126212](#)
 - SAX (Symbolic Aggregate approXimation):
 - [J. Lin, E. J. Keogh, S. Lonardi, and B. Y. chi Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in DMKD, 2003, pp. 2–11.](#)
 - [http://www.cs.ucr.edu/~eamonn/SAX.htm](#)
 - [http://code.google.com/p/jmotif/wiki/SAX](#)
 - [https://github.com/jmotif/sax-vsm_classic](#)

Información en Trenes de Potenciales de Acción (Temporal Coding)

- Supongamos un tren de *spikes* y un Δt , de tal forma que cada incremento de tiempo pueda contener al menos un potencial de acción.
- Codificamos los *spikes* como 1's y la ausencia como 0's.



Representation of spike train as binary string with time resolution Δt .



Información en Trenes de Potenciales de Acción (*Temporal Coding*)

- Suponemos que los *spikes* son independientes.
- El *firing rate*, r , determina el número de unos y ceros que hay en un tren de disparos de longitud T .
- Podemos calcular el número de posibles patrones de *spikes* con un número fijo de éstos en una secuencia de longitud T segundos.
- Así, la entropía máxima es logaritmo en base 2 de ese número (problema combinatorio).

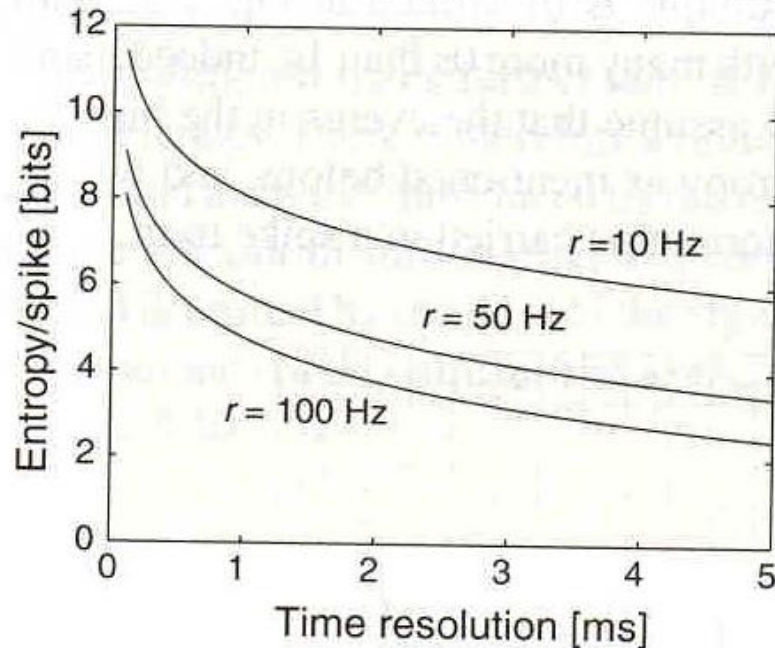


Información en Trenes de Potenciales de Acción (*Temporal Coding*)

- Se puede demostrar que cuando $\Delta t \ll 1/r$, entonces la entropía es: $S \approx Tr \log_2 (e/(r \Delta t))$, donde el factor $Tr=N$ es el número de disparos en una secuencia de longitud T .
(demostrar para casa (5))
- Así, la entropía máxima por disparo viene determinada por $(S/N) \approx \log_2 (e/(r \Delta t))$, donde podemos ver que depende de la precisión y del *firing rate* que elegimos.
- Esta medida nos da la máxima entropía por disparo cuando podemos asignar a cada uno de ellos un único mensaje.

Información en Trenes de Potenciales de Acción (Temporal Coding)

A. Maximum entropy with spike code



B. Maximum entropy with rate code

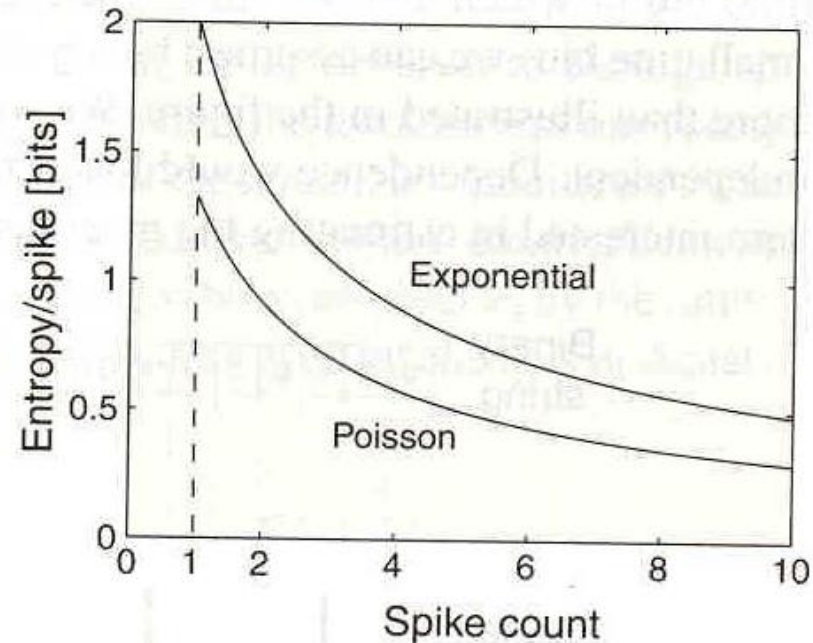


Fig. 5.10 (A) Maximum entropy per spike for spike trains with different firing rates as a function of the size of the time bins, which sets the resolution of the firing times. (B) Maximum entropy per spike for spike trains from which the information is transmitted with a rate code. The two different curves correspond to two different spike statistics of the spike train, a Poisson and an exponential probability of spike counts. Spike trains with exponential spike distributions can convey the maximum information with a rate code for fixed variance.



Información en Trenes de Potenciales de Acción (*Temporal Coding*)

- Lo más remarcable es que la entropía máxima por disparo es siempre mayor que 1.
- La razón es que los intervalos entre disparos es grande y por tanto, la ocurrencia de un *spike* es altamente relevante para el sistema (altamente informativo).
- También vemos que la entropía decrece cuando incrementamos el parámetro Δt , debido a que el número de patrones que se pueden representar es menor, ya que la ocurrencia de un disparo no es tan relevante.



Información en Trenes de Potenciales de Acción (*Temporal Coding*)

- También la entropía decrece cuando aumentamos el *firing rate* en el tren de disparos de longitud T , ya que un disparo individual ya no es tan sorprendente como antes.
- Igualmente, se puede demostrar que la **información decrece** cuando existe **ruido** en el canal de transmisión y cuando hay **correlaciones** entre los disparos.



Información en Trenes de Potenciales de Acción (*Rate Coding*)

- Vamos a aplicar ahora la herramienta de teoría de información a un esquema de codificación *Rate Coding*.
- Recordemos que hasta ahora hemos medido para todos los posibles patrones de disparos, lo que equivale a *Temporal Coding*.
- En *Rate Coding* lo que es relevante es el número de *spikes*, n , contenido en una secuencia de los mismos de longitud T .
- Así, la entropía de observar N *spikes* de media en un tiempo T es $S(N;T) = -\sum_n P(n) \log_2 P(n)$, donde **$P(n)$ es la probabilidad de observar n disparos en un intervalo T .**



Información en Trenes de Potenciales de Acción (*Rate Coding*)

- Supongamos que la secuencia de disparos se puede aproximar por un **proceso estocástico poissoniano**. Es decir, la $P(n)$ viene determinada por una distribución de Poisson: $P(n) = (N^n e^{-N}) / n!$, donde el parámetro de la distribución, N , es el promedio de disparos que nosotros esperamos, $N = rT$, es decir el valor esperado de la variable aleatoria n (número de disparos) es $E(n) = \sum_n P(n)n = N$.
- Se puede deducir que $S(N;T) \approx 0.5 \log_2 N - 0.5 \log_2 (2\pi)$, y si este valor lo dividimos por el número de disparos, volvemos a obtener la entropía por disparo.



Información en Trenes de Potenciales de Acción (*Rate Coding*)

- La entropía por disparo en este caso es más pequeña.
- El tiempo sobre el que se calcula la entropía depende del *firing rate*, i.e. $S(N;T(N))$. Así, por ejemplo, para obtener un *spike count* de 2 tenemos que calcular la entropía sobre un T de 40ms si la neurona dispara con un *firing rate* medio de 50 Hz ($t_f=1/50=0.02$, $0.02*2=40\text{ms}$, dos disparos en 40ms).
- Podríamos preguntarnos cuál es la **distribución** de probabilidad, dada una media y dispersión fijas, que **maximiza la entropía**.



Información en Trenes de Potenciales de Acción (*Rate Coding*)

- La respuesta es una distribución exponencial:
 $P(n) \propto e^{-\lambda n}$ con *firing rate* $r=1/(T(\exp(\lambda)-1))$, y así obtenemos:
$$S(N;T)=\log_2(1+N)+N \log_2(1+(1/N))$$
- De nuevo, si dividimos por N tenemos entropía por disparo.
- Vemos que con esta distribución la entropía aumenta.
- Este esquema de codificación **no** es capaz de llevar tanta información como el *temporal coding*. Pero es razonable (2 bits/spike de información), en el límite donde un *spike* aislado lleva toda la información.



Información en Trenes de Potenciales de Acción (*Dificultades*)

- En la realidad, medir entropías en datos reales es muy *costoso*, porque tenemos que estimar la distribución de probabilidad. Necesitamos muchos datos para construir histogramas.
- Otro problema es que la información depende del logaritmo de las probabilidades. Así, **sucesos muy poco probables contribuyen de forma fuerte a la información**, pero como son muy improbables hay que esperar a que salgan.
- Cuando normalizamos los histogramas de frecuencias (probabilidad) entonces tendemos a **sobrestimar** el contenido de información, ya que perdemos los eventos muy improbables.



Información en Trenes de Potenciales de Acción (*Neuronas Aisladas*)

- Para medir la información mutua entre dos neuronas, tenemos que calcular la distribución de disparos de cada unidad por separado y la distribución de probabilidad conjunta.
- Sin embargo, es más simple medir una sola neurona y la distribución de respuesta de esta neurona respecto a diferentes estímulos ($P(y|s)$).
- Así podemos calcular: $I(s) = \sum_y P(y|s) \log_2 (P(y|s)/P(y))$
- Esta medida es casi una medida de información, ya que hemos utilizado probabilidad condicional (se pueden perder las propiedades aditivas).
- Definimos esta cantidad como la **sorpesa** de la neurona dependiente del estímulo.

Información en Trenes de Potenciales de Acción (*Neuronas Aisladas*)

- El promedio de esta cantidad sobre $P(s)$ se puede demostrar que es la información mutua de dos señales (el estímulo y la neurona)

$$MI(y,s) = \sum_s \sum_y P(s) P(y|s) \log_2 (P(y|s)/P(y))$$

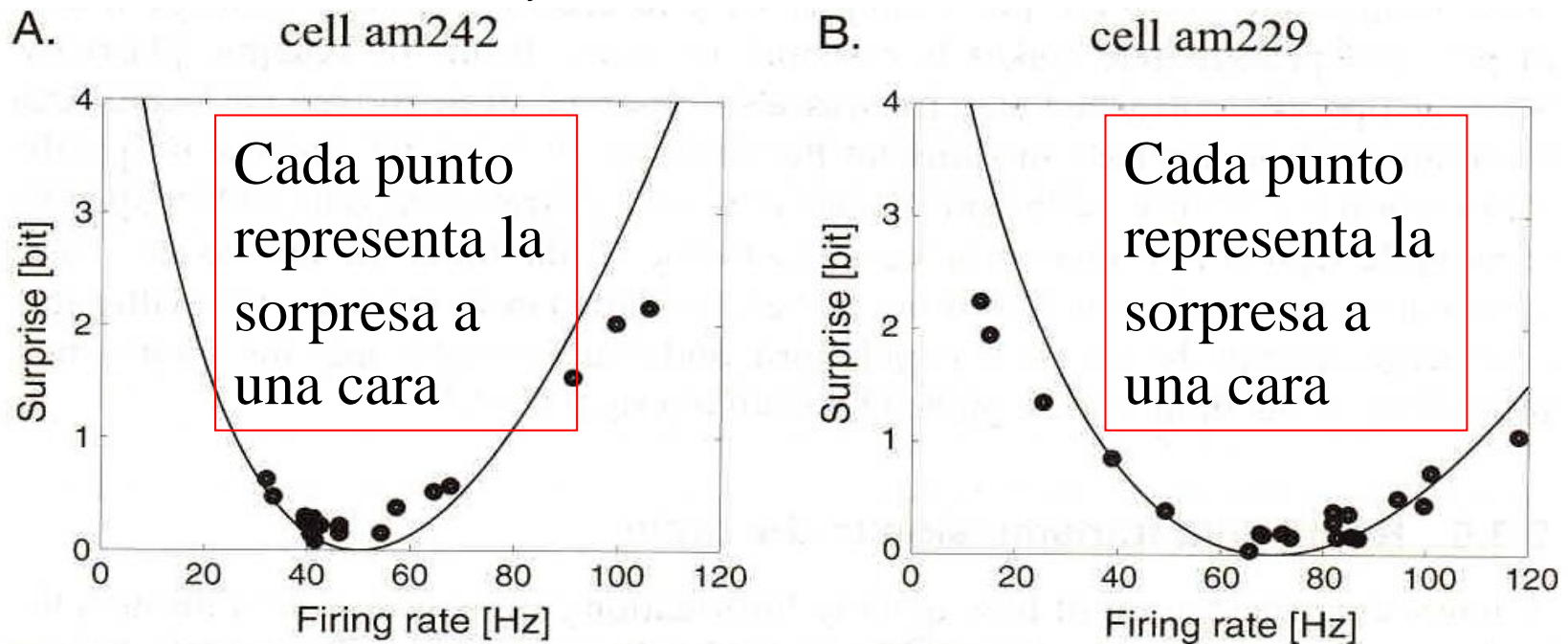


Fig. 5.11 Stimulus-dependent surprise for two cells in the inferior temporal cortex that respond to faces. The curve is 5 times the universal expectation for small time windows (eqn 5.30) [data from Rolls, Treves, Tovee, and Panzeri, *J. Comp. Neuro.* 4: 309–33 (1997)].



Información en Trenes de Potenciales de Acción (*Neuronas Aisladas*)

- En la figura A el *fring rate* promedio fue de 54Hz.
- Se puede observar que las respuestas alrededor de ese valor promedio tienen asociado valores bajos de sorpresa.
- Sin embargo, desviaciones fuertes tienen valores de sorpresa sobre 2 bits por disparo.
- No obstante, el promedio de la sorpresa en todas las respuestas de la neurona, y frente a 20 caras, es más o menos de 0.55 bits (estimación de información mutua $MI(s,y)$ con 20 estímulos, caras, para s).
- No es un valor muy grande, lo que nos sugiere que no tiene por qué utilizarse *Temporal Coding* para que estas neuronas trabajen.



Información en Trenes de Potenciales de Acción (*Neuronas Aisladas*)

- En la figura B sucede más o menos lo mismo. Sin embargo, el promedio es algo mayor, y tiene un valor alto de la sorpresa del estímulo para valores bajos de la tasa de disparo.
- Los valores de la sorpresa fueron calculados para 500ms, durante la presentación de las caras.
- Hay muchos estudios que demuestran que la información de un estímulo transmitida se realiza en intervalos de tiempo suficientemente cortos, de tal forma que más de un disparo pueda ocurrir en esos intervalos.
- Así, no hay mucha libertad para esquemas de codificación complejos (como *temporal coding*) y se ha ajustado la sorpresa del estímulo mediante una curva universal (ver la curva de ajuste en las figuras).



Información en Trenes de Potenciales de Acción (*Neuronas Aisladas*)

- Esta curva es $I^s(\Delta t \ll 1) = r^s \log_2(r^s/r) + ((r-r^s)/\ln 2)$, donde r^s es la tasa de respuesta específica para el estímulo s , y r es el promedio de la tasa de disparo sobre todos los estímulos.
- En definitiva, la información puede ser transmitida en tiempos cortos mediante el esquema de *rate coding*.



Información en Trenes de Potenciales de Acción (Neuronas Aisladas)

- Evidentemente, **esto no quiere decir que la precisión de los disparos no sea importante en el cerebro.**
- Así, por ejemplo, la emisión de disparos de una neurona que temporalmente están muy cerca de neuronas conectadas a ésta, puede **estimular ondas de actividad** en el cerebro.
- Esto es debido a lo que ya vimos sobre los integradores con pérdida.
- También la **precisión es fundamental** para la **plasticidad sináptica**. Concretamente, los tiempos de disparo presinápticos y postsinápticos.
- Igualmente se observa **correlación** de disparos entre **hemisferios** del cerebro.



Información en Trenes de Potenciales de Acción: Transmisión rápida de los datos en el cerebro.

- Un ejemplo de cómo se transmiten los datos en el cerebro fue demostrado por Simon Thorpe y colegas.
- Demostraron que los humanos son capaces de discriminar la presencia de animales o coches en escenas que se presentan en un intervalo corto de tiempo (20 ms).
- Se presenta a continuación el porcentaje de respuestas correctas cuando el sujeto liberaba un botón e identificaba un animal en una imagen compleja, que se le presentaba durante 20 ms. Se hizo para 15 sujetos.

Información en Trenes de Potenciales de Acción: Transmisión rápida de los datos en el cerebro.

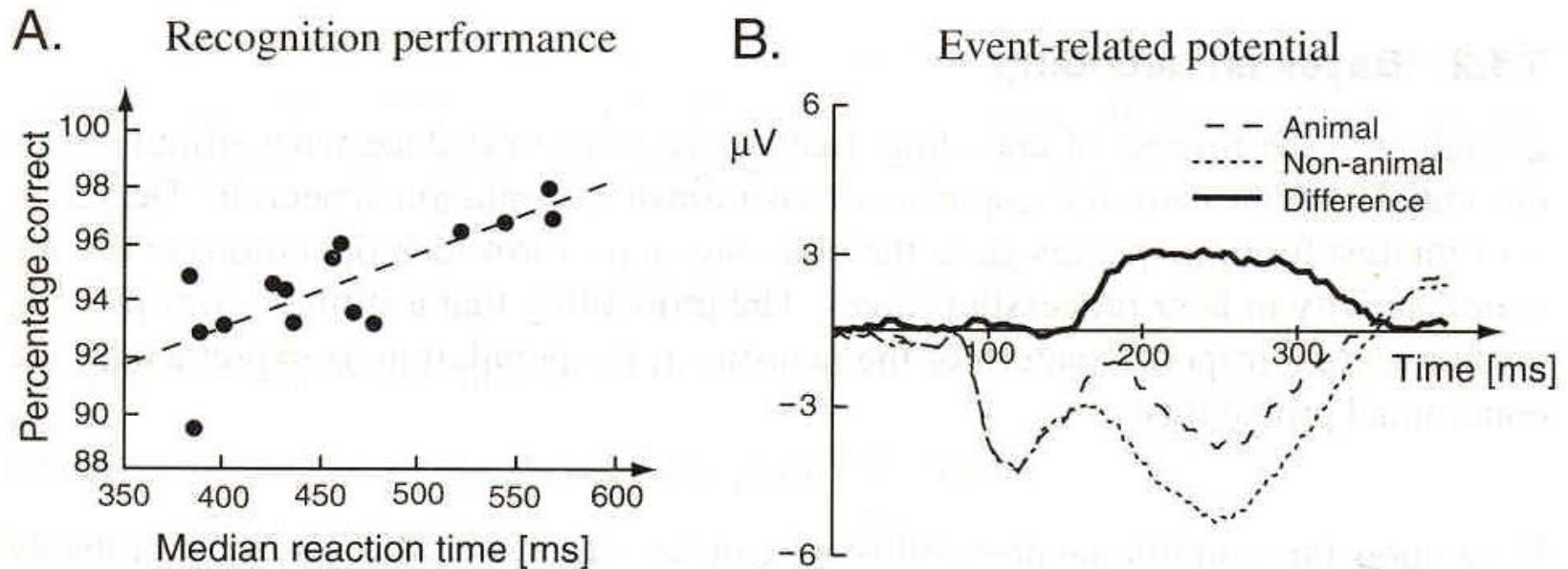


Fig. 5.12 (A) Recognition performance of 15 subjects who had to identify the presence of an animal in a visual scene (presented for only 20 ms) versus their mean reaction time. (B) Event-related potential averaged over frontal electrodes across the 15 subjects [redrawn from Thorpe, Fize, and Marlow, *Nature* 381: 520–22 (1996)].



Información en Trenes de Potenciales de Acción: Transmisión rápida de los datos en el cerebro.

- En la figura se puede observar la gran habilidad para reconocer objetos en esos tiempos tan cortos de presentación del estímulo, y en un tiempo de reacción tan escaso.
- También registraron encefalogramas durante los experimentos en el *cortex* frontal.
- En los primeros 150 ms no hay diferencia, pero a partir de aquí el *cortex* frontal cambia su actividad cuando distingue un animal en la escena.
- Esto es importante, ya que se sabe que este tipo de procesamiento de información tiene que pasar a lo largo de varias capas en diferentes áreas cerebrales. Y se puede estimar que cada neurona tiene que pasar la información en órdenes de tiempo de 10 a 20 ms.



Codificación y Decodificación en Poblaciones Neuronales.

- Es razonable pensar que la información sobre el mundo externo se represente o se procese de una forma distribuida en toda la completitud del cerebro.
- Anteriormente hemos considerado varios esquemas de codificación.
- Aquí vamos a considerar la distribución espacial de la información en una población de neuronas.
- Utilizaremos un marco probabilístico debido a la mencionada variabilidad neuronal de los sistemas biológicos.



Codificación y Decodificación en Poblaciones Neuronales.

- **Codificación:** Es la forma específica en la que el estímulo es representado en la neurona o población neuronal.
- **Decodificación:** Es la manera en que el mensaje que está distribuido en la población puede ser interpretado de acuerdo a la respuesta de la neurona o población neuronal.



Codificación en Poblaciones Neuronales.

- Podemos considerar la codificación de un patrón de estímulos en función de una probabilidad de respuesta de la población neuronal.

$$P(r|s)=P(r_1^s, r_2^s, r_3^s,|s),$$

donde s es cierto estímulo y r_i^s es la respuesta específica de la neurona i respecto al estímulo s .

- Principalmente, se consideran para las respuestas frecuencias o tasas de disparos, ya que son fáciles de medir.
- Sin embargo, otro tipo de medida para la respuesta neuronal es completamente válido.
- P puede ser una distribución de probabilidad continua o discreta.